Mapping of the Proposed CTEP Simplified Disease Classification Terms to Concepts in NCI Thesaurus

Recently CTEP developed and circulated a new classification called the CTEP Simplified Disease Classification, to solicit comments. An evaluation of the proposed classification was conducted at COH to verify if the classification fits into an existing vocabulary/ontology resource. Since SNOMED CT is not a cancer specific vocabulary, the SNOMED searches did not yield suitable results in many cases. We also attempted to map the CTEP classification terms to the NCI Thesaurus. The NCI Thesaurus is an excellent vocabulary resource whose contents are oriented towards cancer, but also includes many other areas, e.g., anatomy, healthcare administrative areas, molecular genetics. It is a true ontology based on an expression language called Description Logic. Each concept can have relationships with other concepts. This has been exploited in the Thesaurus for creating rich sets of relationships in several cancer areas e.g., gastroenterology and breast.
The mapping exercise showed that the CTEP classification is sound and has significant correlation with the NCI thesaurus:

Total terms in CTEP classification	223
The CTEP classification terms with an **exact match** with an NCI Thesaurus concept	183 (82%)
The closest NCI Thesaurus concept **broader** than the term in CTEP classification	21 (9.4%)
The closest NCI Thesaurus concept **narrower** than the term in CTEP classification	11 (4.9%)
The closest NCI Thesaurus concept only **approximately matches** the term in CTEP classification	4 (1.8%)
There is **no appropriate** NCI Thesaurus concept that is even close to the CTEP term in CTEP classification	4 (1.8%)

"Broader" and "narrower" indicate that both the concepts are part of the same concept inheritance hierarchy. "Broader" means that the NCI concept can subsume the CTEP term but has other concepts in its scope too. "Narrower" NCI concept means that that the concept that the CTEP term denotes can subsume the NCI thesaurus concept. Approximately matching NCI concept means that both the concepts are similar and often are children of the same broader concept and may have an overlap.

The discrepancies in the classification and the thesaurus may be resolved by making changes to the CTEP classification to conform to the NCI concepts. It may even be worthwhile to use the NCI thesaurus concept names for all the entities in the CTEP classification. If a semantic conflict is revealed in the changing of CTEP classification, recommendation may be made for changes in the NCI thesaurus. The NCI Thesaurus is administratively and technically amenable to such changes.

Conclusions:

1. The NCI thesaurus can be very for many of the purposes in caBIG. Also, there appears to be a consensus within the caBIG VCDE space about value of the NCI thesaurus.
2. The proposed CTEP classification is good and with little tweaking can become completely compliant with the NCI Thesaurus. The CTEP classification can be used for doing much of the reporting - this will require mapping of the classification to other vocabularies also, like SNOMED CT, ICD-03 and ICD9 CM. The Category and Sub-category entities of the classification should also be mapped to the NCI Thesaurus, which was not done in the present exercise. The CTEP also has mappings of the classification to MedDRA, legacy CTEP etc., which they may be willing to share.

The mappings can be seen in the accompanying spreadsheet.